Lec 15 MATH 393

The strong, core assumption which gave us inference,

$$\vec{\mathcal{E}} \sim \mathcal{N}_n(\vec{0}_n, \sigma^2 I_n) \iff \mathcal{E}_1, \ldots, \mathcal{E}_n \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

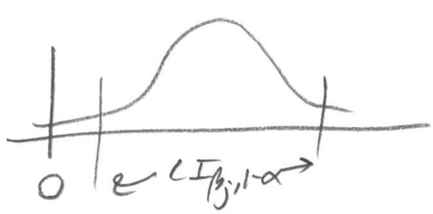
Can be wrong in many ways. "Robust Regression" are methods that provide valid inference under weaker assumptions.

$\boxed{\text{Scenario \#1}}$

$\mathcal{E}_1, \ldots, \mathcal{E}_n \overset{ind}{\sim}$ then bootstrapping would work.

Let X_b be the design matrix constructed from n indices that were sampled with replacement from $\{1, 2, \ldots, n\}$. Let $\vec{y}_b$ be the responses. Let $\vec{b}_b = (X_b^T X_b)^{-1} X_b^T \vec{y}_b$

If you're interested in testing $H_0: \beta_j = 0$ at level α, then examine all b_{b_j} for many bootstrap samples and create a $CI_{\beta_j, 1-\alpha}$. If $0 \notin CI_{\beta_j, 1-\alpha}$ $\Rightarrow$ Reject H_0



Scenario #2

$\mathcal{E}_1, \ldots, \mathcal{E}_n \overset{iid}{\sim}$ unknown DGP s.t. $E[\mathcal{E}] = 0$ and $Var[\mathcal{E}] = \sigma^2$

mean-centered and homoskedastic but not necessarily normal.

Scenario #2

very common scenario!!

$\mathcal{E}_1, \ldots, \mathcal{E}_n \overset{iid}{\sim}$ s.t. $E[\mathcal{E}_i] = 0$ and $Var[\mathcal{E}_i] = \sigma^2$, mean-centered, homoskedastic errors but not necessarily normal

Under this scenario,

$E[\vec{B}] = \vec{\beta}$ and $Var[\vec{B}] = \sigma^2 (X^\top X)^{-1}$ but $\vec{B}$ is not $\sim N(\vec{\beta}, \sigma^2 (X^\top X)^{-1})$

so we don't get the t-test nor F-tests. What can we do?

It turns out that we can show

$$\vec{B} \,\dot{\sim}\, N(\vec{\beta}, \sigma^2 (X^\top X)^{-1}) \Rightarrow \frac{1}{\sigma} \vec{B} \,\dot{\sim}\, N(\vec{\beta}, (X^\top X)^{-1})$$

This is difficult to prove, so I'll omit it.

We still don't know σ^2. So we use Slutsky's theorem to get a Wald test

$$\frac{1}{s_e} \vec{B} \,\dot{\sim}\, N(\vec{B}, (X^\top X)^{-1})$$ which gives you approx testing

$\Rightarrow$ You can basically use regression output results if $n >> p+1$ as the $T_{n-(p+1)} \approx N(0,1)$

But we don't get the F-tests. How do we run tests such as $H_0: \vec{\beta}_S = \vec{0}_{|S|}$

where $S \subseteq \{0, 1, \ldots, p\}$

Recall two facts from 340. let $S \subseteq \{1, 2, \dots, k\}$

submatrix

$$\vec{X} \sim N_k(\vec{\mu}, \Sigma) \Rightarrow \vec{X}_S \sim N_{|S|}(\vec{\mu}_S, \Sigma_{S \times S})$$ proved using jacobians

$$\Rightarrow (\vec{X} - \vec{\mu})^T \Sigma^{-1} (\vec{X} - \vec{\mu}) \sim \chi^2_k$$ Mahalanobis distance

let's apply those facts here:

$$\overset{\text{Fact 1}}{\Longrightarrow} \frac{1}{s_e} \vec{b}_S \overset{\cdot}{\sim} N\left(\vec{\beta}_S, (X^T X)^{-1}_{S \times S}\right) \overset{\text{Fact 2}}{\Longrightarrow} \frac{1}{s_e^2} (\vec{b}_S - \vec{\beta}_S)^T \left((X^T X)^{-1}_{S \times S}\right)^{-1} (\vec{b}_S - \vec{\beta}_S) \overset{\cdot}{\sim} \chi^2_{|S|}$$

which can be used to do all F-tests. For example, to do the omnibus F, $S = \{1, \dots, p\}$, $H_0 : \vec{\beta}_S = \vec{0}_p$

which goes by a Wald test

if $\frac{1}{s_e^2} \vec{b}_S^T \left((X^T X)^{-1}_{S \times S}\right)^{-1} \vec{b}_S > \chi^2_{p, 1-\alpha} \Rightarrow$ Reject H_0.

which is an approx test.

Scenario #3

$\varepsilon_i \overset{ind}{\sim} N(0, \sigma_i^2)$. Here, the variances are different (heteroskedastic).

$\Updownarrow$

$\vec{\varepsilon} \sim N_n(\vec{0}_n, D)$ s.t. $D = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots & \sigma_n^2 \end{bmatrix}$

In this case

$$\vec{b} = (X^T X)^{-1} X^T \vec{Y} = (X^T X)^{-1} X^T (X \vec{\beta} + \vec{\varepsilon}) = \vec{\beta} + (X^T X)^{-1} X^T \vec{\varepsilon}$$

$$\overset{340}{\Longrightarrow} \vec{B} \sim N\left(\vec{\beta}, (X^T X)^{-1} X^T \operatorname{Var}(\vec{\varepsilon}) \left((X^T X)^{-1} X^T\right)^T\right)$$

$$= N\left(\vec{\beta}, (X^T X)^{-1} X^T D X (X^T X)^{-1}\right)$$

But we don't know D since σ_i^2's are unknown.

Note that $\forall i$ $E[\varepsilon_i^2] = \sigma^2$ since $E(\varepsilon_i) = 0$

so E_i^2 is likely a good estimator for σ^2

let $\hat{D} = \begin{bmatrix} E_1^2 & & 0 \\ & E_2^2 & \\ & & \ddots \\ 0 & & E_n^2 \end{bmatrix}$

In 1980, White proved that $(X^\top X)^{-1} X^\top \hat{D} X (X^\top X)^{-1} \xrightarrow{p} (X^\top X)^{-1} X^\top D X (X^\top X)^{-1}$

Slutsky

$\Rightarrow \vec{B} \overset{\cdot}{\sim} N\left(\vec{\beta}, (X^\top X)^{-1} X^\top \hat{D} X (X^\top X)^{-1}\right)$

little abuse of notation!

Heteroskedastic - Consistent Error Regression Inference

Note: other $\hat{D}$'s have been proposed. No time to elaborate!

$\boxed{\text{Scenario \#4}}$

Same as scenario #3 except normality ~~not~~ assumed $\Rightarrow$ Same as Scenario #3 except more approximate.

How to do F-tests? Wald test as in scenario #2.

Recommendation: use Bootstrap! The core assumption is always false!

Other robustness methods $\boxed{\text{NOT COVERED}}$

$\vec{\mathcal{E}} \sim N(\vec{0}, \Sigma) \iff \vec{Y} \sim N(X\vec{\beta}, \Sigma)$

$\Rightarrow$ Maybe try to transform $\vec{y}$ to be normal?

Box-Cox transformations, Tukey-Mosteller transforms

Other lm tests. If $p+1 = 2$ e.g. $H_0: \beta_1 = 17\beta_2 \Rightarrow [1 \; {-17}]\vec{\beta} = 0$

Multiple testing: Scheffé's, so much more!!!